Journal of Network and Information Security
10 (2) 2022, 15-20
http://www.publishingindia.com/jnis

Object Detection Assistance for Visually Impaired People - A Brief Review

Aayushi Balothiya[1*], Amol Paliwal[2] and Nisha Rathi[2]
[1]Acropolis Institute of Technology and Research, Indore, Madhya Pradesh, India.
Email: aabalothiya001@gmail.com
[2]Acropolis Institute of Technology and Research, Indore, Madhya Pradesh, India.
*Corresponding Author

Abstract: One of our most vital senses is sight because it accounts for 80% of all information, we take in. This paper is a systematic literature review about object detection for visually impaired people, in which the work done covers 20+ research papers related to this topic and provides answers to different research questions to provide the reader with an easily readable summarized overview of the papers related to this topic. The YOLO approach has many advantages over other methods of object detection. When compared to other algorithms, the YOLO algorithmic system fully scans a picture by anticipating binding boxes via a network of communication and the potential for these boxes. Additionally, the YOLO algorithm's architecture and functioning are presented in this study.

Keywords: Convolution, Neural network, YOLO.

I. Introduction

Vision is one of the most important sensors that individuals use to interact with things in the real world. People with vision can swiftly assess their surroundings by looking about, determining what is nearby, how far away it is, and how to interact with it. It is simple to view people going about their regular activities because they can see everything around them well, including any impediments they may encounter and other people, making it simple to engage with these things.

However, because of their everyday tasks and activities, visually impaired persons must work hard to survive in the outside world. It is a well-known statistic that there are 285 million visually challenged persons worldwide, or roughly 20 percent of India's population, and that having a visual impairment interferes with many daily activities. Knowing their surroundings and being able to comprehend their experiences is therefore among the most crucial things for blind people.

The main causes of visual impairments, which affect millions of individuals worldwide today, are age-related issues, diabetes, and accidents. One of the top ten disabilities for both men and women is blindness. Blind and visually impaired people face difficulties such as loss of flexibility and default, job loss, inability to do assignments, inability to make independent purchases, and time and effort needed to complete simple tasks.

Now that new technology is available, efforts are being made to enable the blind lead regular lives. Although new goods are being developed quickly, some browsing issues arise because of any social engagement. NAVI, VOICE, and TVS Tyflos systems were additional technologies used before Yolo. The TVS (Tactile Vision System) and Tyflos systems have a device in the user's stomach that converts depth into vibration. NAVI is based on Visual Navigation Assistance Disability that distinguishes between background and obstacle. VOICE system converts image into audio so that the user can see the obstacle. These processes are cumbersome and inefficient [26].

II. Existing Literature and Recent Work

In 2018, Juan Du published a review of the literature titled Understanding of Object Detection Based on CNN Family and YOLO. After describing the backdrop of CNNs, this paper discusses the You Only Look Once (YOLO) algorithm. The two variations of YOLO that are addressed in this study are YOLO V1 and V2. The algorithms, traits, and layers of YOLO are listed. By contrasting YOLO and Faster R-CNN in terms of speed, complexity, accuracy, and cost, their benefits and drawbacks may be shown. Finally, the outlook for the future of CNN is provided as a summary of YOLO and Faster R-CNN [12].

In 2019, Jonathan Shen, Ruoming Pang, and Ron J. Weiss published Object Detection and Distance Estimation Tool for Blind People using Convolutional Methods with Stereovision. In this paper, a system that can convey details about nearby objects is created. To make it simpler for blind people to use, this system can also determines the distance of an object that has been identified using a camera that is paired with glares. They can definitely detect objects in their environment and develop their skill and ability with the aid of this technology. This system produces real-time video as visual data using a camera as the primary sensor, which functions similarly to the human eye. CNN [13] is the algorithm in use.

In 2019, Avanti Dorle, Piyush Pimplikar, Pranit Bagmar, and Atharva Rajkuvar released their "Object Recognition App for Visually Impaired." This paper discusses an Android application that combines many techniques to create one that will not only assist those who are blind in recognising items around in real-time but will also help them as soon as possible by providing an auditory output. The SSD (Single Shot Detector) technique is used for object detection and recognition. This method has shown to be quicker than other alternatives and provides extremely accurate results for real-time objects. The programme also uses the Android TextToSpeech and Android Tensorflow APIs to provide audio output [14].

In 2019, Geethapriya S., N. Duraimurugan, and S. P. Chokkalingam published Real-Time Object Detection using Yolo.

The goal of the work is to thoroughly examine the YOLO method. Additionally, it compares various object identification and distance estimate techniques in terms of speed, accuracy, and complexity. By employing a neural network to forecast the bounding boxes and the class probabilities for these boxes, the YOLO method thoroughly examines the image and recognizes it more quickly than previous algorithms to research various YOLO applications [15].

Sunit Vaidya, Niti Shah, and Nisha Shah developed Real-Time Object Detection for Visually Impaired People in 2020. The suggested application in this work uses machine learning algorithms and image processing to identify things in real-time using a camera and alert a blind person to the object's location via auditory output. The major goal of the proposed work is to improve the world for persons who are visually impaired by offering the best performance outcomes, good precision, and a practical option. Because of its quick speed and high accuracy, the YOLO algorithm is used [16].

Roshan Rajwani, Paresh Kalinani, Dinesh Purswani Indu Dokare and Deesha Ramchandani developed a Proposed System on Object Detection for People with Visual Impairments. Along with several supporting libraries, the programme mostly uses Android. An image of the surroundings will be taken using the camera on an Android smartphone and stored in the operating system's memory. Libraries like OpenCV and the Google Cloud Vision API will be used to process this image. Google Cloud is used by Google Cloud Vision API. Through the internet, the image is uploaded to the cloud. The supplied image is compared to millions of other photos using the COCO dataset. The procedure is finished, and the things in the image are recognized. The user receives information via an audio output about the recognized things that are present in his environment.

CICERONE - A Real-Time Object Detection for Visually Impaired People was developed by Therese Yamuna Mahesh, Parvathy S. S., Shibin Thomas, Shilpa Rachel Thomas and Thomas Sebastian. The goal of this project is to develop a low-cost technology that will allow someone with poor vision to live normally and independently in each setting, such as our home or place of employment. Using a walking stick with a Raspberry Pi attached all items utilized for daily activities can be discovered. The procedure is depicted in Fig. 1's flowchart. Fig. 1 shows the system's suggested flowchart. The photographs are obtained with a Raspberry Pi 4 model B. Using rclone, the captured photos are uploaded to the Google Drive. For object identification and categorization, YOLO network is employed. gTTS (Google Text to Speech module) is used to convert the test to speech for the detected class and send it to the users' earbuds [18].

Object Detection using Machine Learning for Visually Impaired People by Venkata Naresh Mandhala, Debnath Bhattacharyya, Vamsi B. and Thirupathi Rao N. considering that general algorithms are typically implemented in OpenCV, known as the well-known computer vision library, the architecture has been bought out. These methods have also been employed in earlier iterations of these object identification theories. The most recent apps that are built using these emerging technologies are less precise and useful. As a result, under some conditions, these old algorithms could not satisfy its requirements for evaluating its performance and job efficiency [19].

III. Evolution of CNN for Object Detection

- *Origin (Late 1980s to Early 1990s):* LeNet-5, created in 1998 by LeCun *et al.*, was the first widely used CNN. It has been in the works for around ten years. To get handwritten numerals was the goal. It is acknowledged as driving CNN's most successful area of deep learning research and development. At ATMs, banks began utilising it.
- *Standing (Early 2000s):* At the time, nobody knew how CNN worked internally. Additionally, there was no data set for different image types like Microsoft COCO or Google OpenImages. Consequently, a lot of CNNs concentrated only on character recognition (OCR). CNNs also require longer computation times and cost more to operate. CNN was outperformed by the machine learning model Support Vector Machine (SVM).
- *Revival (2006-2011):* Ranzato *et al.,* demonstrated in their study that feature extraction is much improved when a big compound approach is utilised in place of the formerly-used small sample algorithm [1]. The GPUs were already being used by researchers to accelerate CNN training. Simultaneously, NVIDIA unveiled the CUDA platform, which enables consistent processing and speeds up CNN training and validation [2]. The research was renewed by this. Another barrier was removed when Stanford University introduced Pattern Analysis, Statistical Modeling and Computational Learning Visual Object Classes (PASCAL VOC), a sizable photo database.
- *Rise (2012–2013):* AlexNet represented a significant improvement in CNN accuracy. In the 2012 ILSVRC

competition, it only had an error rate of 15.3 percent. The error rate for the network in second place was 26.2% [3]. Therefore, AlexNet outperformed every other network at the time by a wide margin of 10.8%. AlexNet successfully implemented "deep learning" by using a total of 8 layers [3] to attain this accuracy. Greater computational power was needed for this, but advancements in GPU technology made it possible. One of the most important articles to ever appear on CNN, along with LeNet, is AlexNet.

- *New Developments in Architecture (2014–2020):* In 2014 [4], the well-known and popular VGG architecture was created. Like many previous VGG-based systems, RCNN introduced the notion that objects are concentrated areas of the image, hence the name "region-based CNN." [4]. In the years that followed, Fast RCNN [5] and Faster RCNN [7]—improved versions of RCNN—were released. Both improved computing efficiency while retaining the RCNN's famed accuracy. A similar VGG-based technology known as Single Shot MultiBox Detector (SSD) was created in 2016 [8]. You Only Look Once (YOLO), a different algorithm built on the DarkNet architecture, was originally introduced in 2016 [8]. Its third version was launched in 2018 [6], and it is now in active development.

IV. YOLO Algorithm for Object Detection

Several modes are available for object detection, including R-CNN Speed, Retina-Net, and Single-Shot MultiBox Detector (SSD). Even though these techniques have addressed the issues of data limits and modelling in object acquisition, they are unable to identify objects using a single algorithm operation. Due to the YOLO algorithm's superior performance versus the discovery methodologies, it has grown in popularity. Convolutional neural networks (CNN) are used by the YOLO method to find objects in real-time. The technique just needs one forward transmission across the neural network to detect objects, as the name would imply. This indicates that a single algorithm is used to anticipate the entire image. CNN is utilized to simultaneously forecast different class opportunities and bond boxes. There are several different variations of the YOLO algorithm. Other common ones include the YOLO and the smaller YOLO-v3.

Fig. 1: YOLO's Development in Recent Years

The following justifications make the YOLO algorithm crucial:

Speed: Because this algorithm can predict objects in real-time, it increases the speed of detection.

High Accuracy: The YOLO prediction method yields precise findings with few background mistakes.

Learning Capabilities: The algorithm has great learning capabilities that allow it to pick up on object representations and use them for object detection.

YOLO's architecture, which was modelled after the GoogleNet design, consists of 24 convolutional layers in total, with 2 completely linked layers at the very end.

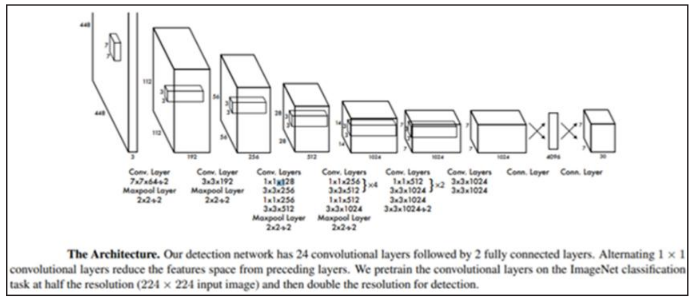

Fig. 2: YOLO Architecture

The following three methods are used by the YOLO algorithm to function. Intersection Over Union (IOU), residual blocks, and bounding box regression [28].

The image is divided into different grids in residual blocks. S x S are the dimensions of each grid. There are numerous grid cells with equal dimensions in the following image. Any objects that enter a grid cell will be detected there. For instance, if the center of an object comes within a particular grid cell, that cell will oversee detecting it (Fig. 3).

Fig. 3: Residual Blocks: An Input Image is Divided into Grids

A bounding box in bounding box regression is an outline that draws attention to an object in a picture. The following characteristics are present in each bounding box in the image: Height (bh), Width (bw), and Class (for example, person, car, traffic light, etc.) The letter c stands for this. Boxing in the middle (bx, by). To forecast the height, width, center, and class of objects, YOLO use a single bounding box regression. It reflects the likelihood of an object existing in the bounding box in the image above (Fig. 4).

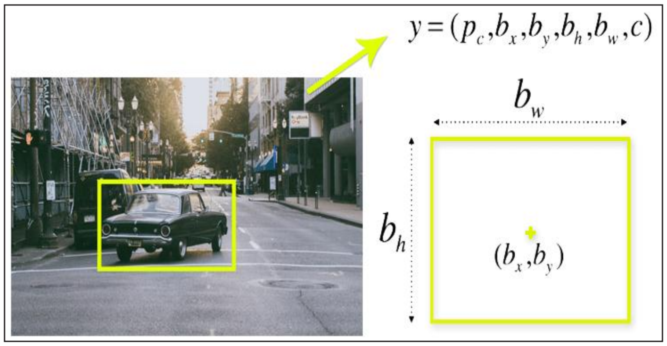

Fig. 4: Bounding Box: Represented by a Yellow Outline

The cross-section of the union (IOU) is a requirement in the acquisition of anything in Intersection Over Union (IOU), which explains how the boxes are separated. To ensure a smooth exit, YOLO uses the IOU. Combination boxes and their confidence scores are predicted by each cell grid. If the predictable binding box resembles a genuine box, then the IOU is equivalent to 1. Binding boxes that do not fit inside an actual box are removed by this machine. A straightforward illustration of an IOU's operation is shown in the figure below (Fig. 5). Two bounding boxes, one in green and the other in blue, can be seen in the image. The green box is the actual box, while the blue box is the anticipated box. YOLO ensures that the two bounding boxes are equal.

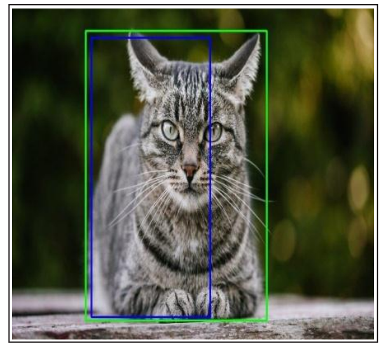

Fig. 5: Intersection Over Union (IOU)

A Fusion of the Three Methods: The three procedures are applied to create the final detection results, as seen in the accompanying image. Grid cells are first used to divide up the image. Each cell grid forecasts B-binding boxes and rates the confidence of its points. Each item's odds of finding a stage are predicted by cells. A car, a dog, and a bicycle, for instance, fall into at least three different categories of items. A single convolutional neural network is used to make all predictions simultaneously. The predictable binding boxes are equal to the real object boxes thanks to road crossings. This requirement gets rid of extra binding boxes that don't meet material requirements (such as length and width). The final discovery will include special binding boxes that are perfectly suited to the artefacts. The bicycle is enclosed by a yellow bracelet, while the car is encompassed by a pink binding box (Fig. 6). A blue combo box is used to highlight the dog [29].

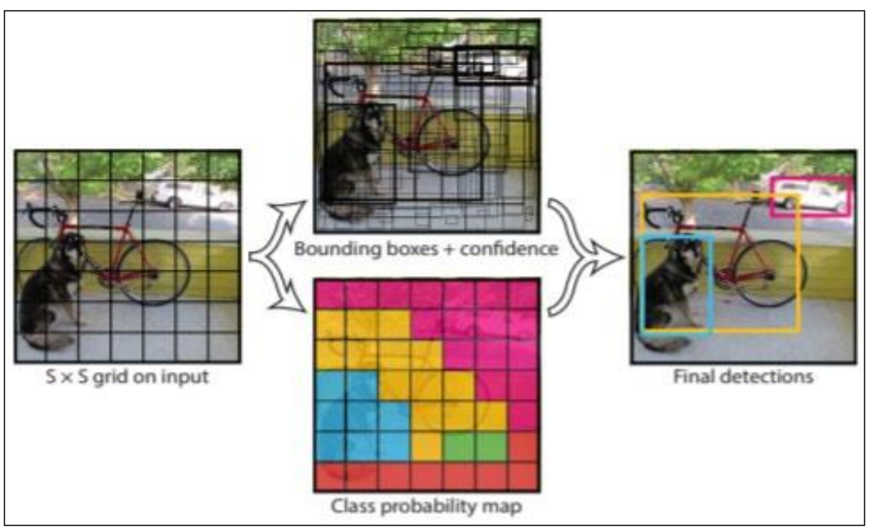

Fig. 6: Combination of Residual Blocks, Bounding Box and Intersection Over Union (IOU) Methods

V. Faster RCNN

Accuracy of this algorithm comes at the cost of time complexity. It is significantly slower than the likes of YOLO.

Despite improvements over RCNN and Fast RCNN, it still requires multiple passes over a single image unlike YOLO.

FRCNN has many components—the convolutional network, Regions of Interest (ROI) pooling layer and Region Proposal Network (RPN). Any of these can serve as a bottleneck for the others.

VI. Comparative Study

Year	Authors Name	Title	Object Detection Method	Computation Time	Accuracy
2015	Hanen Jabnoun, Faouzi Benzarti, Hamid Amiri	Object Detection and Identification for Blind People in Video Scene [23]	Scale Invariant Features Transform (SIFT) and the Speed Up Robust Features (SURF) algorithms	Moderate	High
2020	V. Balaji, S. Kanaga Suba Raja, C. J. Raman, S. Priyadarshini, S. Priyanka, S. P. Salai Kamakathai	Real-Time Object Detection for visually Impaired using Open [24]	Python 3.5, Caffe model framework, OpenCV, Numpy 1.14	Moderate to High	High
2018	Roshan Rajwani, Dinesh Purswani, Paresh Kalinani, Deesha Ramchandani, Indu Dokare	Proposed System on Object Detection for Visually Impaired People [17]	OpenCV and Google Cloud Vision API	Low	Low
2020	Therese Yamuna Mahesh, Parvathy S. S., Shibin Thomas, Shilpa Rachel Thomas and Thomas Sebastian	CICERONE - A Real-Time Object Detection for Visually Impaired People [18]	YOLO algorithm	High	High
2020	Venkata Naresh Mandhala, Debnath Bhattacharyya, Vamsi B., Thirupathi Rao N.	Object Detection using Machine Learning for Visually Impaired People [19]	Classification/clustering techniques	High	Low to Moderate
2020	Pooja Maid, Omkar Thorat, Sarita Deshpande	Object Detection for Blind Users [25]	Tensorflow algorithm and RCNN	Moderate	Moderate
2020	K. Vijlyakumar, K. Ajitha, A. Alexia, M. K. Vijlyakumar, K. Ajitha, A. Alexia, M. Hemalashmi and S. Madhumitha	Object Detection for Visually Impaired People Object Detection for Visually Impaired People using SSD Algorithm [26]	SSD algorithm, Monodepth algorithm	Moderate to High	High

VII. Conclusion

This paper briefly addresses the various object detection algorithms for people with visual impairments, including the CNN family, SSD, and YOLO. On the basis of many characteristics, including speed, cost, accuracy, and complexity, various algorithms are compared. In comparison to CNNs and SSD, YOLO has more sophisticated applications in practice. A unified object detection model is YOLO. It is simple to construct and may be trained directly on entire images. The YOLO algorithm is fast when compared to other algorithms. However, the YOLO algorithm's primary flaw is that it is less accurate than other algorithms [20].

References

[1] M. A. Ranzato, F. J. Huang, Y. L. Boureau, and Y. LeCun, "Unsupervised learning of invariant feature hierarchies with applications to object recognition," in 2007 *IEEE Conference on Computer Vision and Pattern Recognition*, IEEE, 2007, p. 1-8.

[2] J. Nickolls, I. Buck, M. Garland, and K. Skadron, "Scalable parallel programming with CUDA: Is CUDA the parallel programming model that application developers have been waiting for?," *Queue*, vol. 6, no. 2, pp. 40-53, 2008.

[3] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "Imagenet classification with deep convolutional neural networks," *Adv Neural Inf Process Syst*., vol. 25, pp. 1097-1105, 2012.

[4] R. Girshick, J. Donahue, T. Darrell, and J. Malik, "Rich feature hierarchies for accurate object detection and semantic segmentation," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2014, pp. 580-587.

[5] R. Girshick, "Fast R-CNN," in *Proceedings of the IEEE International Conference on Computer Vision*, 2015, p. 14408.

[6] J. Redmon, and A. Farhadi, "Yolov3: An incremental improvement," 2018, arXiv preprint arXiv:1804.02767.

[7] S. Ren, K. He, R. Girshick, and J. Sun, "Faster R-CNN: Towards real-time object detection with region proposal networks," *IEEE Trans Pattern Anal Mach Intell*., vol. 39, no. 6, pp. 1137-1149, 2016.

[8] J. Redmon, S. Divvala, R. Girshick, and A. Farhadi, "You only look once: Unified, real-time object detection,"

in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2016, pp. 779-788.

[9] T. Ahmad, Y. Ma, M. Yahya, B. Ahmad, and S. Nazir, "Object detection through a modified YOLO neural network," *Scientific Programming*, vol. 2020, 2020, Art. id 8403262.

[10] W. Liu, D. Anguelov, D. Erhan, C. Szegedy, S. Reed, C. Y. Fu, and A. C. Berg, "SSD: Single shot multibox detector," in *European Conference on Computer Vision*, Cham: Springer, 2016, pp. 21-37.

[11] M. Gang, J. Zhiguo, L. Zhengyi, Z. Haopeng, and Z. Danpei, "Full-viewpoint 3D space object recognition based on kernel locality preserving projections," *Chinese Journal of Aeronautics*, vol. 23, no. 5, pp. 563-572, 2010.

[12] J. Du, "Understanding object detetction based on CNN family and YOLO," *J. Phys.: Conf. Ser.*, vol. 1004, 2018, Art. no. 012029.

[13] J. Shen, R. Pang, R. J. Weiss, M. Schuster, N. Jaitly, Z. Yang, "Object detection and distance estimation tool for blind people using convolutional methods with stereovision," 2019.

[14] A. Dorle, P. Pimplikar, P. Bagmar, and A. Rajkuvar, "Object recognition app for visually impaired," 2019.

[15] Geethapriya S., N. Duraimurugan, and S. P. Chokkalingam, "Real-time object detection with Yolo," 2019.

[16] S. Vaidya, N. Shah, and N. Shah, "Real-time object detection for visually challenged people," 2020.

[17] R. Rajwani, D. Purswani, P. Kalinani, D. Ramchandani, and I. Dokare, "Proposed system on object detection for visually impaired people," *International Journal of Information Technology (IJIT)*, vol. 4, no. 1, Mar.-Apr. 2018.

[18] T. Y. Mahesh, S. S. Parvathy, S. Thomas, S. R. Thomas, and T. Sebastian, "CICERONE - A real-time object detection for visually impaired people," *IOP Conference Series: Materials Science and Engineering,* vol. 1085, p. 012006, Feb. 2021, doi: https://doi.org/10.1088/1757-899x/1085/1/012006.

[19] V. N. Mandhala, D. Bhattacharyya, Vamsi B., and Thirupathi Rao N., "Object detection using machine learning for visually impaired people," *International Journal of Current Research and Review*, vol. 12, no. 20, pp. 157-167, 2020, doi: http://dx.doi.org/10.31782/IJCRR.2020.122032.

[20] P. Kasture, A. Kumkar, Y. Jagtap, A. Tangade, and A. Pole, "A survey on "object detection algorithms for visually impaired people"," *International Journal of Advanced Research in Science, Communication and Technology (IJARSCT)*, vol. 3, no. 2, Mar. 2021

[21] S. Srivastava, A. V. Divekar, C. Anilkumar, I. Naik, V. Kulkarni, and V. Pattabiraman, "Comparative analysis of deep learning image detection algorithms," *J Big Data*, vol. 8, 2021, Art. no. 66, doi: https://doi.org/10.1186/s40537-021-00434-w.

[22] Z.-Q. Zhao, P. Zheng, S.-T. Xu, and X. Wu, "Object detection with deep learning: A review", *IEEE Transactions on Neural Networks and Learning Systems*, vol. 99, pp. 1-21, Jan. 2019, doi: 10.1109/TNNLS.2018.2876865.

[23] H. Jabnoun, F. Benzarti, and H. Amiri, "Object detection and identification for blind people in video scene," *2015 15th International Conference on Intelligent Systems Design and Applications (ISDA)*, 2015, pp. 363-367, doi: 10.1109/ISDA.2015.7489256.

[24] J. Nasreen, W. Arif, A. A. Shaikh, Y. Muhammad, and M. Abdullah, "Object detection and narrator for visually impaired people," *2019 IEEE 6th International Conference on Engineering Technologies and Applied Sciences (ICETAS)*, 2019, pp. 1-4, doi: 10.1109/ICETAS48360.2019.9117405.

[25] P. Maid, O. Thorat, and S. Deshpande, "Object detection for blind user's," *International Research Journal of Engineering and Technology (IRJET)*, e-ISSN: 2395-0056, vol. 7, no. 6, Jun. 2020.

[26] K. Vijlyakumar, K. Ajitha, A. Alexia, M. Hemalashmi, and S. Madhumitha, "Object detection for visually impaired people using SSD algorithm," *2020 International Conference on System, Computation*, *Automation and Networking (ICSCAN)*, 2020, pp. 1-7, doi: 10.1109/ICSCAN49426.2020.9262344.

[27] R. Deepa, E. Tamilselvan, E. S. Abrar, and S. Sampath "Comparison of Yolo, SSD, Faster RCNN for real-time tennis ball tracking for action decision networks," *International Conference on Advances in Computing and Communication Engineering (ICACCE)*, 2019.

[28] G. Karimi, "Introduction to YOLO algorithm for object detection," EngEd Community, Apr. 15, 2021.

[29] M. Maithani, "Guide to Yolov5 for real-time object detection," Developers Corner, Dec. 19, 2020.